

AYUSH SINHA

(332) 209-9928 | ayush.sinha@columbia.edu | [GitHub](#) | [LinkedIn](#)

EDUCATION

Columbia University

MS in Data Science, GPA: 3.82/4.0

Coursework: Probability, Statistical Inference, Algorithms, Deep Learning, Machine Learning, Cloud Computing

TA: Algorithms for Data Science (Fall 2022)

New York, NY

Sep 2021 - Dec 2022

Manipal Institute of Technology

BTech in Information Technology, Minor in Computational Math, GPA: 9.24/10.00

Coursework: Data Structures, Operating Systems, Databases, Embedded Systems, Distributed Systems, Computer Networks

Manipal, IN

Jun 2017 - Jul 2021

SKILLS

- Languages/Libraries: C++, Python, Java, JavaScript, HTML, CSS, SQL, Matplotlib, NumPy
- Framework/Tools: Sklearn, TensorFlow, Keras, Linux, Git, Docker, Spark, Hadoop, EC2, RDS, S3, EB, BigQuery, Excel, Flask

WORK EXPERIENCE

Columbia Experimental Gravity Group

Research Assistant

New York, NY

Jun 2022 – Present

- Working on developing the Low-Latency Algorithm for Multi-Messenger Astrophysics (LLAMA) **search pipeline** to receive and identify **real-time** astrophysical signals and generate skymap json files with low latency
- Implemented a new hop client listener to keep receiving real-time gravitational wave and neutrino signal through **Kafka data stream**, as well as parsing, storing, analyzing the real-time astrophysics event message and generating significance calculation files

Columbia Business School

Research Assistant

New York, NY

Jul 2022 - Aug 2022

- Extracted and analyzed S&P 500 companies' sustainability reports; Generated Sentence Embeddings using **ClimateBERT**
- Ranked relevant sentences utilizing cosine distance and ran array jobs on Columbia HPC for ~150 reports
- Fine-tuned ClimateBERT for sequence classification on imbalanced dataset using Hugging Face, reaching accuracy of 90%

Altair Engineering

Deep Learning Intern

Troy, Michigan

May 2022 - Aug 2022

- Reviewed custom implementation of Graphical U-Net to replace simulations; Introduced **OOP design** to components
- Designed Novel Optimized **pooling algorithm** and feature mapping operation in **Neural Network** achieving speedup factor of 96
- Experimented with **sparse inputs** to reduce memory consumption of Model below 250 GB while training

Dell Technologies

Business Intelligence Intern

Bangalore, IN

Feb 2021 - Jun 2021

- Extracted data using SQL; Integrated using Python and Excel; Performed data exploration and feature selection
- Utilized **logistic regression** and **Random forests** to determine customer as promoter or detractor with accuracy > 60%
- Analyzed ~30 features from model outputs to draw business interpretations and communicated findings to > 10 teams

PROJECTS

Radiology Report Generation ([link](#))

Sep 2022 - Dec 2022

- Supported **Accenture AI team** in utilizing multimodal learning from Radiology Reports and Chest X-rays for report generation
- Compared the performance with SOTA and investigated model results through experimentation
- Created the inference pipeline of the model and replaced **ResNet** with **EfficientNet** observing 25% increase in BLEU score

Student Management Application

Sep 2022 - Dec 2022

- Developed a scalable full stack application to support **CRUD** operations, pagination, login authentication, notifications, etc
- Devised application frontend using Angular and deployed using **AWS CloudFront**; Conforming API to be RESTful
- Built **Microservices** using Flask and deployed using AWS EC2 and EB and deployed DB using AWS RDS

Spectral Representations for Convolutional Neural Networks ([link](#))

Nov 2021 - Dec 2021

- Coded custom spectral pooling layer, spectral parametrization of filters and frequency dropout for CNN using TensorFlow
- Visualized information preservation during spectral pooling and convergence speed in spectral convolutions
- Achieved faster convergence of CNN during training with factor of 2.2-5.1 based on architecture and a competitive pooling method

Dell Technologies: Competitive Intelligence, WiCi Platform

Jun 2020 - Sep 2020

- Implemented **web crawlers** to automate aggregation of data from multiple competitor websites using BeautifulSoup and Selenium
- Transformed data for efficient use and built a classification pipeline for various news articles with accuracy > 90%
- Optimized web crawling using multithreading saving time by 80% and summarized news articles using **TextRank**

Kochi Metro Data Analysis: Won Manipal Coursera Competition ([link](#))

Jun 2019 - Jul 2019

- Did EDA of Kochi Metro data, found correlations in Metro traffic flows and Pre-processed ridership data inspired by SOTA papers
- Applied **PCA** and built a customized **KNN** Regression model for predicting short term passenger ridership with MAPE ~ 15